

Jaakko Leppiniemi

SISÄILMAN LAADUN HEIKKENEMISEN EHKÄISEMINEN DATA-ANALYTIIKAN MENETELMIN

Tekniikan ja luonnontieteiden tiedekunta
Kandidaatintyö
Joulukuu 2019

TIIVISTELMÄ

Jaakko Leppiniemi: Sisäilman laadun heikkenemisen ehkäiseminen data-analytiikan menetelmin

Use of data-analytics in preventing indoor air quality decline

Kandidaatintyö

Tampereen yliopisto

Tietojohtaminen

Joulukuu 2019

Sisäilman laadun heikkeneminen on ollut ongelma siitä asti, kun ihmiset ovat eläneet sisätiloissa. Sisäilma saastuu erilaisista yhdisteistä, kuten orgaanisista yhdisteistä, pienhiukkasista ja liiallisesta kosteudesta. Sisäilman saastuminen johtaa ihmisten hyvinvoinnin heikkenemiseen, opiskelun tai työnteon tehokkuuden heikkenemiseen ja näiden kautta taloudellisiin tappioihin. Rakennuksia joudutaan purkamaan tai korjaamaan liian usein, mikä maksaa valtiolle turhaa rahaa. Ilmastointilaitteita käytetään joko liikaa tai liian vähän, jolloin energiankulutus kasvaa. Nämä ongelmat voitaisiin ehkäistä seuraamalla ja kehittämällä ilmanlaatua.

Aluksi täytyy mitata pienhiukkasten konsentraatioita ja havainnoimalla sisäilmaa. Mittareista tulee paljon ja erilaista dataa, eli big dataa. Dataa täytyy analysoida, jotta saadaan tietoa sisäilman tilasta. Analyysin tekemiseen on useita työkaluja, kuten koneoppisen avulla luotavat ennustavat mallit. Näiden mallien avulla ilmanvaihtojärjestelmät voidaan ohjata optimaaliseen suoritukseen, jolloin säästyy energiaa, terveyttä ja rahaa.

Tutkittua analytiikkaa ilmanlaadun kontekstissa voidaan esittää kuusiportainen prosessi tiedon hankintaan. Ensin tunnistetaan tietotarpeet, jonka jälkeen mitataan dataa, jota sittemmin analysoidaan. Sen jälkeen arvioidaan prosessia ja tuloksia, jotta saadaan tarkkaa tietoa, jota voidaan käyttää päätöksenteon tukena. Tutkimustuloksia voidaan soveltaa esimerkiksi kunnan palveluissa tai yrityksen liiketoiminnassa.

Avainsanat: sisäilman laatu, data-analytiikka, päätöksenteko

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

ALKUSANAT

Tämä tutkimus on tehty kandidaatintyönä Tampereen yliopiston tietojohdamisen tutkinto-ohjelmaan syyslukukaudella 2019. Tutkimuksen aihetta valitessa juteltiin edellisen kesän työnantajani Metosinin henkilöiden ja heidän yhteistyökumppanin Loopshoren toimitusjohtajan kanssa. Loopshore on ilmanlaadun mittauksiin ja datan analysointiin keskittyvä yritys, jonka projekti oli mielestäni kiinnostava. Data-analytiikka liittyy tietojohdamiseen ja on tärkeä taito tulevaisuuden työelämässä, joten työ päätettiin tehdä ilmanlaadun mittauksista saatavan datan analysoinnista.

Haluan kiittää ohjaajaani ja kandidaatintyöni tarkastajaa Miikka Palvalinia, jonka kanssa käydyistä keskusteluista sain uusia näkökulmia työni tekemiseen. Haluan myös kiittää Loopshorea ja sen toimitusjohtajaa Janne Edgreniä, joka antoi idean aiheeseen. Kiitos myös Pyry Collanderille, Arttu Kukkoselle ja kaikille muille, jotka auttoivat minua kandidaatintyön tekemisessä. Lisäksi haluan kiittää saman pienryhmän opiskelijoita vertaistuksesta. Erityisesti haluan kiittää Riku Sundqvistia kandidaatintyöni opponoinnista ja palautteesta.

Tampereella, 3.12.2019

Jaakko Leppiniemi

SISÄLLYSLUETTELO

1. JOHDANTO	1
1.1 Taustaa	1
1.2 Tutkimusongelma ja rajaukset	2
1.3 Tutkimuksen rakenne	2
2. TUTKIMUSMENETELMÄN JA -AINEISTON ESITTELY	4
2.1 Tutkimusmenetelmät	4
2.2 Tutkimusaineiston esittely	5
3. SISÄILMAN LAADUN HEIKKENEMINEN	7
3.1 Sisäilman laadun heikkenemisen syyt	7
3.2 Sisäilman laadun heikkenemisen terveysvaikutukset	9
3.3 Sisäilman laadun heikkenemisen taloudelliset vaikutukset	10
4. DATAN KÄSITTELY	12
4.1 Analytiikkaprosessi	12
4.2 Koneoppiminen	16
5. SISÄILMAN LAADUN KÄSITTELY DATA-ANALYTIIKAN MENETELMIN	17
5.1 Datan käsittely ja analysointi	17
5.2 Tiedolla johtaminen	19
6. YHTEENVETO	21
6.1 Tulosten esittely ja arviointi	21
6.2 Tutkimuksen merkityksellisyys ja lisätutkimusten tarve	22
LÄHTEET	24

1. JOHDANTO

Johdantoluvun tehtävä on tutustuttaa lukija aiheeseen. Luvussa kerrotaan tutkimuksen aiheen taustaa sekä rajataan tutkimusongelma. Lopuksi kerrotaan vielä kandidaatintyön rakenteesta.

1.1 Taustaa

Huono sisäilman laatu on tunnistettu viiden suurimman riskin joukkoon ympäristön vaikutuksessa ihmisten terveydelle. Se voi johtua esimerkiksi radonista, formaldehydistä, homeesta, pölystä sekä erilaisista orgaanisista yhdisteistä. Sisäilman laatu on tärkeä tunnistaa terveydelliseksi vaikuttajaksi, sillä ihmiset saattavat viettää jopa 85% ajastaan sisätiloissa. (Oualet et al. 2018) Suomessa kosteudesta johtuva home on melko yleistä ja on tärkeää levittää tietoa kosteuden aiheuttaman homeen vaikutuksesta sisäilmaan ja sitä kautta ihmisten terveyteen.

Hometta syntyy kosteuden ja lämpötilan yhteisvaikutuksesta (Hoskins 2003). Kosteutta voidaan poistaa rakennuksista hyvällä ilmanvaih dolla ja rakennussuunnittelulla, mutta se on monimutkainen prosessi, jota ei tässä työssä käsitellä. Suomessa ilmasto vaihtelee paljon kostean ja kuivan sekä kylmän ja lämpimän välillä, mikä vaikeuttaa rakennussuunnittelua. Suomessa sisäilmaongelmia voi siis esiintyä lähes kaikkialla, kuten kouluissa, päiväkodeissa, toimistoissa, sairaaloissa ja kasarmeilla. Lähtökohtaisesti siis sisäilmaongelmat ovat vakavia ja yleisiä, minkä vuoksi aiheen tutkiminen on tärkeää. Ulkoseinissä piilevä kosteus on kuitenkin suhteellisen helppoa ja edullista mitata esimerkiksi lämpötilaerojen avulla.

Yhdysvalloissa Berkeleyn yliopiston laboratoriossa tehdyn tutkimuksen mukaan sisäilman huonoksi havaittu laatu vaikuttaa työtehokkuuteen (Berkeley Lab 2019b). Havaittu laatu tarkoittaa ilmanlaatua, jonka ihminen aistii itse, huolimatta mittaustuloksista ja yhdisteiden konsentraatioista. Samaan tutkimukseen liittyvissä kyselyissä työtehokkuus parani prosentin jokaista kymmenen prosentin nousua kohden havaitussa sisäilman laadussa. (Berkeley Lab 2019b) Langerin et al. (2017) toteuttamassa tutkimuksessa selvitettiin sisäilman havaitun laadun ja ilmansaasteiden suhdetta. Tutkimuksessa ilmanlaatua mitattiin erilaisilla mittareilla sekä asukkaiden ulkopuolisten tarkastajien vierailuilla ja havainnoinnilla. Tutkimuksessa selvisi, että ulkopuoliset tarkastajat havaitsivat ilman

huonommaksi kuin asukas itse. Mittarit näyttivät myös enemmän saasteita ilmassa, mitä huonommaksi sisäilma havaittiin. (Langer et al. 2017) Tästä voidaan siis päätellä, että huononevan ilmanlaadun voi havaita mittareilla jo ennen kuin se alkaa vaikuttaa suorituskykyyn. Tämän ilmiön tutkiminen onkin keskeinen osa tämän työn kulkua.

1.2 Tutkimusongelma ja rajaukset

Tämän kandidaatintyön keskeiset tutkimuskysymykset käsittelevät tapaa, jolla mittauksista saatavaa dataa voidaan hyödyntää ihmisten eduksi. Tässä työssä tutkitaan vastauksia yhteen olennaiseen pääkysymykseen:

- Miten ilmanlaadun heikkeneminen voidaan estää ennen kuin se vaikuttaa ihmisten hyvinvointiin negatiivisesti?

Ilmanlaadun heikkenemisen estäminen vaatii tietynlaisia päätöksiä, joiden tekemiseen tarvitaan tietoa ongelmasta. Tätä ongelmaa tutkitaan datalähtöisesti ja työssä pyritään löytämään analyttisiä menetelmiä mittareista saatavan datan tulkitsemiseksi. Dataa ja-
lostamalla saadaan tietoa, jota voidaan käyttää päätöksenteon tukena esimerkiksi rakennus- ja materiaaliteknisessä suunnittelussa.

Ihmisten hyvinvointi on tärkein prioriteetti ja ensimmäinen ongelma, jota halutaan tutkia lisää. Kandidaatintyö ei kuitenkaan ole tarpeeksi laaja koko asian käsittelemiseen, jolloin aihetta täytyy rajata enemmän. Mikä hyvinvoinnin osa-alue kaipaa tutkimista? Mikä alue ilmanlaadusta kaipaa tutkimista? Kirjoittajan oman mielenkiinnon sekä työn järkevän etenemisen kannalta tutkimuksen kohde rajataan sisäilmaan ja sen laadun vaikutuksiin. Eri-
tyisesti työssä kiinnostaa sisäilman laadun heikkeneminen. Miksi se heikkenee ja miten se voidaan estää? Näitä kysymyksiä tarkastellaan työn edetessä.

1.3 Tutkimuksen rakenne

Tutkimus etenee suoraviivaisesti aiheen taustasta ongelmien ja kysymysten kautta itse tutkimukseen. Tutkimuksessa kerrotaan myös, miten aihetta tutkitaan ja mitä näkökulmia tarkastellaan. Aluksi on johdantoluku, jossa avataan tutkimusongelmaa ja motiiveja sen tutkimiseen. Johdannon jälkeen toisessa luvussa kerrotaan tutkimusmenetelmistä ja esitellään aineisto. Kolmannessa luvussa perehdytään tarkemmin ilmanlaadun ja sen mittaamisen teoriaan, jonka jälkeen neljännessä luvussa esitellään löydettyjä prosesseja ja malleja datan muuntamiseksi tiedoksi. Tutkimuksessa ei perehdytä mallien teknisyyteen syvemmin, mutta joitain löydettyjä datan analysointitapoja esitellään. Viidennessä luvussa yhdistetään kaksi edellistä käsittelykappaletta ja sidotaan analytiikkaprosessi il-

manlaadun mittauksien kontekstiin. Luvussa esitellään myös prosessien tulosten esittämisen keinoja. Lopuksi viimeisessä luvussa on tutkimuksen yhteenveto ja tulosten esittely sekä tutkimuksessa käytettyjen lähteiden luettelo.

2. TUTKIMUSMENETELMÄN JA -AINEISTON ESITTELY

Tässä luvussa kerrotaan yleisesti tutkimuksen toteutuksesta. Aluksi kerrotaan tutkimusmenetelmistä, jonka jälkeen esitellään tutkimuksessa käytetty aineisto.

2.1 Tutkimusmenetelmät

Kandidaatintyö on kirjallisuuskatsaus. Työhön kuului aiheeseen liittyvän kirjallisuuden etsiminen ja tutkiminen. Aineistoja analysoitiin ja vertailtiin työn jokaisessa kohdassa ja tietoa koottiin eri lähteistä yhdeksi kokonaisuudeksi. Kirjallisuuskatsaus tehtiin noudattaen Finkin (2005 Salminen 2011 mukaan) mallia, jonka vaiheet voi katsoa taulukosta 1.

Taulukko 1. Finkin malli (2005, Salminen 2011 mukaan).

Vaihe	Toimenpiteet
Tutkimuskysymyksen asettaminen	Tutkimusongelma ja -kysymys selkeästi esitettynä
Bibliografian ja tietokantojen ja sivustojen valinta	Mitä tietokantoja käytetään hakuun. Esimerkiksi Andor
Hakutermien valinta	Termien valinta siten, että hakutulokset vastaavat tutkimuskysymykseen
Käytännön seulan asettaminen	Hakutulosten seulonta kielen ja vuoden mukaan
Metodologisen seulan asettaminen	Artikkeleiden ja tutkimusten tieteellisen laadun arviointi
Katsauksen suorittaminen	Standardoitu muoto, jonka mukaan artikkeleista kerätään tietoa
Synteesin tekeminen	Tämänhetkisen tiedon raportointi, tutkimustarpeen osoittaminen ja löydösten selittäminen laadullisesti ja tilastollisesti

Tietoa etsittiin Tampereen yliopiston kirjaston tietokannoista, kuten Andorista ja Tunlibistä. Myös Scopus ja Web of Science olivat hyviä tietokantoja tiedon etsimiseen. Työtä aloittaessa huomattiin, että aiheeseen liittyvä kirjallisuus oli yleensä tieteellisiä artikkeleita, joihin liittyi jokin tutkimus. Tietoa oli saatavilla myös suomeksi, mutta englannin kielellä oli paljon enemmän ja kattavampia tekstejä tarjolla, joten työssä keskityttiin englanninkieliseen kirjallisuuteen. Hakuja tehtiin lausekkeilla, joissa yhdisteltiin ilmanlaatuun ja teknologiaan liittyviä termejä AND- ja OR-operaattoreilla. Hakulausekkeena oli esimerkiksi "perceived indoor air quality" AND "forecasting". Tällaisella lausekkeella saatiin

tuloksia, jotka käsittelivät sekä ennustamista että havaittua sisäilman laatua. Jos haluttiin hakea tietoa joko havaitusta sisäilman laadusta tai ennustamisesta, valittiin AND-operaattorin sijaan OR-operaattori.

Taulukko 2. Eri hakulausekkeilla saadut tulokset eri tietokannoista.

Hakulauseke	Andor	Web of Science	Scopus	Tunilib
"indoor air quality" OR IAQ	129877	8305	11322	119
"perceived indoor air quality" OR "perceived IAQ"	448	84	104	1
("indoor air quality" OR IAQ) AND data	58766	1819	2425	68
("indoor air quality" OR IAQ) AND forecasting	1838	23	125	0
("indoor air quality" OR IAQ) AND "machine learning"	458	22	31	0

Hauissa pyrittiin myös etsimään tietoa synonyymeillä. Huomattiin, että "indoor air quality" on usein lyhennetty "IAQ", joten ilmanlaatuun liittyvät haut tehtiin molemmilla termeillä. Kuten taulukosta 1 huomataan, Andor on selkeästi monipuolisin tietokanta. Web of Science ja Scopus vaikuttavat hyvinkin samankaltaisilta, mutta Tunilibistä ei tässä työssä ole paljoakaan hyötyä.

Hakujen tuloksista etsittiin otsikon perusteella aiheeseen sopiva teksti, joka luettiin nopeasti. Jos teksti oli työhön sopiva, lähde tallennettiin Zotero-lähteidenhallintapalveluun, jotta se olisi helppo löytää myöhemminkin. Tarkoituksena oli etsiä paljon lähteitä, jotta olisi varaa valita mitä lähteitä työssä käytetään. Tieteellisissä julkaisuissa olleet artikkelit ja tutkimukset olivat hieman kirjoja sopivampia työn aiheeseen. Hakuja löytyi yleisemmillä termeillä useita tuhansia, mutta hakuja tarkentamalla ja tietokannasta riippuen, tulokset saatiin rajattua alle tuhanteen ja vielä tarkemmin jopa alle sataan. Liian tarkkaa lausekkeesta ei kannattanut tehdä, sillä joitain hyödyllisiä lähteitä saattoi silloin jäädä löytämättä. Työssä pyrittiin etsimään aiheesta myös vastakkaisia tutkimuksia, jotta aiheesta voisi tehdä mahdollisimman kattava ja puolueeton tutkimus.

2.2 Tutkimusaineiston esittely

Tutkimuksessa käytetty kirjallisuus käsitteli koulujen, kotien ja toimistojen sisäilmaa. Työhön on valittu siis paikkoja, joissa ihmiset viettävät paljon aikaansa. Ilmanlaadusta ja sen heikkenemisen vaikutuksesta terveyteen löytyi paljon artikkeleita. Hyvä lähde sisäilman laadusta ja sen saastumisen terveysvaikutuksista oli John Hoskinsin kirjoittama artikkeli, joka on julkaistu Indoor and Built Environment -julkaisussa vuonna 2003. Siinä on ker-

rottu monipuolisesti, mistä huono sisäilman laatu johtuu ja miksi se on vakava riski terveydelle. (Hoskins 2003) Hoskinsin tutkimusta oli käytetty lähteenä myös useassa muussa tätä työtä varten luetussa artikkelissa, jotka liittyivät sisäilman laatuun. Langer et al. (2017) kertovat sisäilman havaitusta laadusta ja sen suhteesta mitattuun laatuun Ranskassa tehdyssä tutkimuksessaan. Sisäilman laatu on ollut Suomessakin paljon esillä. Finell et al. (2018) esittelevät tutkimuksessaan sisäilman vaikutusta kouluissa oppilaisiin. Tutkimuksessa esitetään myös psykososiaalisten tekijöiden vaikutusta huonosta sisäilmasta johtuviin ongelmiin. (Finell et al. 2018) Huonon sisäilman laadun teorian lisäksi etsittiin lähteitä datan käsittelyyn liittyen. Oualet et al. (2018) kertovat artikkelissaan formaldehydin konsentraation mittauksista saadun datan käsittelystä ja pitoisuuksien ennustamisesta. Wang et al. (2019) sekä Wei et al. (2019) käsittelivät teksteissä koneoppimista datan analysointiin liittyen. Koneoppiminen on yksi tärkeä työkalu, jolla voidaan tehdä kattavia analyyskejä big dataan liittyen Big data tarkoittaa suuria datamääriä monipuolista dataa, jota kerätään nopealla tahdilla (Crowder & Carbone 2018). Tutkimuksessa käytettiin myös organisaation oppimiseen ja päätöksentekoon liittyviä tekstejä, joissa esitellään yritysjohtoon päätöksentekoprosesseja sekä niihin vaikuttamista.

Lähteitä löytyi paljon. Aihetta voi rajata siis datan hyödyntämiseen päätöksenteon tukena käyttäen tietoteknisiä mallinnusmenetelmiä. Työn voi odottaa tukevan aikaisempia tutkimuksia huonon sisäilman laadun haitoista sekä ratkaisumalleista. Tavoitteena on myös löytää tapoja näiden mallien hyödyntämiseen. Aiheen on todettu olevan tärkeä ja ajankohtainen, sillä kouluissa ja sisätyössä näkyvät huonon sisäilman ongelmat joka päivä. Tutkimukselta toivotaan, että se voi auttaa tämän ongelman tietoisuuden levittämistä.

3. SISÄILMAN LAADUN HEIKKENEMINEN

Tässä luvussa vastaan tutkimuskysymykseen: Miten sisäilman laatu vaikuttaa ihmisten hyvinvointiin. Luvussa kerrotaan teoriaa sisäilman laatuun liittyvistä sairauksista sekä vaikutuksista oppimiseen ja työtehokkuuteen.

3.1 Sisäilman laadun heikkenemisen syyt

Sisäilman laadun heikkeneminen voi tuntua kuumuudelta, kosteudelta ja yleiseltä tunkkaisuudelta, mutta syyt huonon sisäilman taustalla ovat paljon laajemmat (Hoskins 2003). Zhang ja Smith (2003) esittelevät tekstissään huonoon sisäilmaan vaikuttavia tekijöitä, joita syntyy osittain normaalin elämisen seurauksena ja osittain ympäristön vaikutuksesta. Niin kauan kuin ihmisiä on ollut olemassa, heillä on ollut tarve suojalle. Sisätiloissa tehdään paljon ruokaa ja välillä omaa elinympäristöä täytyy myös siivota. Ruokanlaitosta ja siivouksesta kuitenkin irtoaa sisäilmaan epäpuhtauksia, kuten pienhiukkasia, aromaattisia yhdisteitä, haihtuvia orgaanisia yhdisteitä ja aldehydejä. Tupakoinnista sekä polttoaineiden käytöstä tulee ilmaan esimerkiksi hiilimonoksidia, typpi- ja rikkioksideja sekä arsenikkia ja fluoria. Rakennusmateriaaleista irtoaa myös paljon epäpuhtauksia, varsinkin kun talossa tehdään remonttia tai saneerausta. Lyijy ja asbesti ovat vakavia ongelmia, mutta harvinaisempia kuin muut. Kaikkiin ilmansaasteisiin ihminen ei kuitenkaan voi vaikuttaa. Ulkoilman saasteet ovat suuri ongelma isoissa kaupungeissa ja vaikuttavat paljon sisäilmaan etenkin, jos ilmanvaihto ei ole kunnossa. Myös tontilla olevan maaperän koostumus, radonin määrä ja muut biologiset yhdisteet vaikuttavat sisäilman laatuun. (Zhang & Smith 2003)

Useat lähteet (Daisey et al. 2003; Hoskins 2003; Ben-David et al. 2019) väittävät ilmastoinnin olevan tärkeä osa sisäilman laadun parantamista. Sisäilma alkaa nopeasti tuntumaan tunkkaiselta, jos ilmastointi ei toimi kunnolla. Myös joissakin kouluissa ilmastointi sammutetaan illaksi ja yöksi. Syyksi moni epäilee kustannusten pienentämistä. Ilmastoinnin sammuttaminen yön ajaksi pitää saman ilman ja kosteuden sisällä ja rakenteissa, jolloin ilma ei pääse vaihtumaan ja pysyvä ilmankosteus auttaa mikrobeja kasvamaan (Terveiden ja hyvinvoinnin laitos 2019). Ilmanvaihtoon on siis syytä kiinnittää huomiota rakennuksia suunniteltaessa, jotta rakennuksia ei tarvitsisi remontoida niin pian rakentamisen jälkeen. Ben-Davidin et al. (2019) mukaan ilmastointi täytyy suunnitella siten, ettei ulkoilmasta pääse haitallisia yhdisteitä, kuten otsonia tai pienhiukkasia, sisäilmaan. Jyväskylän yliopistossa tehdyssä tutkimuksessa selvitettiin ilmanlaadun tilaa Pirkanmaalla 2000-luvulla ja tutkittiin eri päästöjen lähteiden vaikutuksia ulkoilmaan. Tutkimuksissa

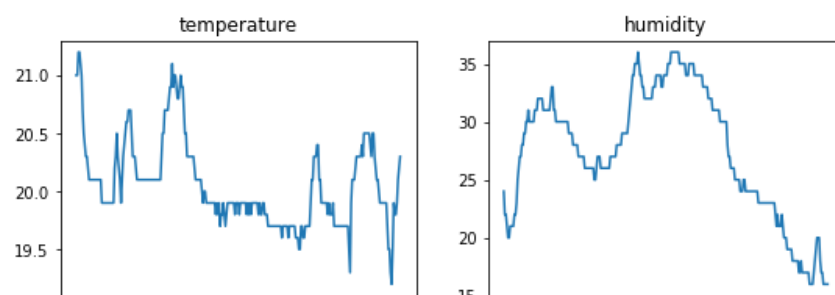
todettiin, että vaikka ilmanlaatu on parantunut, liikenne, energiantuotanto, teollisuus ja katupöly ovat merkittäviä tekijöitä ilmanlaadun heikkenemisen kannalta. (Tampio 2010) Voidaan siis todeta, että huono ulkoilman laatu vaikuttaa myös sisäilmaan, jolloin ilmastoinnin tärkeys korostuu. Köyhemmillä asuinalueilla kehityksistä puhumattakaan ei välttämättä ole varaa rakentaa laadukkaita rakennuksia, joissa olisi hyvin toimiva ilmastointi. Siispä ilmanlaadun tutkiminen erityisesti kustannustehokkaasta näkökulmasta on tärkeää kansanterveyden kannalta, jotta kaikille voidaan taloudellisesta tilanteesta huolimatta tarjota mahdollisuus laadukkaaseen elämään.

Ilmanlaatua täytyy seurata aktiivisesti, jotta se ei pääse heikkenemään. Ilmanlaadun seuraaminen tarkoittaa käytännössä havainnointia säännöllisillä kyselyillä, sekä ilmassa olevien pienhiukkasten ja yhdisteiden konsentraatioiden mittaamista.

	deviceId	timestamp	timestampMS	type	value
0	002EC429	2019-10-30T13:03:41Z	1572440621000	humidity	24.0
1	002EC429	2019-10-30T13:03:41Z	1572440621000	temperature	21.0
2	002EC429	2019-10-30T13:36:01Z	1572442561000	humidity	22.0
3	002EC429	2019-10-30T13:36:01Z	1572442561000	temperature	21.0
4	002EC429	2019-10-30T14:08:20Z	1572444500000	humidity	22.0

Kuva 1. Oikeista mittareista mitattua dataa (Loopshore 2019).

Kyselyistä ja mittauksista saadaan havaintoja ja faktoja eli dataa, kuten kuvasta 1 näkyy. Työtä varten saatiin ilmanlaadun mittaamiseen ja analysointiin erikoistuneelta yritykseltä käytettäväksi havainnollistavaa dataa. Jotta mittauksista saadaan paras hyöty irti, havainnoitua ilmanlaatua täytyy tutkia samaan aikaan ilman eri yhdisteiden pitoisuuksien kanssa. Kuvassa 1 on selkeästi esitettyä mittalaitteen tunnus, mittauksen aikaleima sekä mittauksen tyyppi ja arvo.



Kuva 2. Kuvan 1 datasta saatu viikon ajalta arvot lämpötilalle (vas.) ja kosteudelle (Loopshore 2019).

Mitattua dataa voidaan käsitellä ja analysoida erilaisilla menetelmillä, joita esitellään myöhemmin. Kuvassa 2 on esiteltynä kaksi hyvin yksinkertaista kuvaajaa lämpötilasta ja kosteudesta.

3.2 Sisäilman laadun heikkenemisen terveysvaikutukset

Huono sisäilma aiheuttaa useita erilaisia sairauksia, joiden muodostuminen riippuu ilman epäpuhtauksista. Esimerkiksi tupakansavun mukana tulevat pienhiukkaset ja hiiliyhdisteet aiheuttavat paljon keuhkosairauksia ja radonsäteily syöpää (Hoskins 2003; Zhang & Smith 2003). Ihmisen ei tarvitse polttaa tupakkaa sairastuakseen, vaan esimerkiksi asunnon entisten asukkaiden tupakointi voi näkyä sisäilmassa vielä vuosia ja aiheuttaa jopa syöpää (Zhang & Smith 2003). Siksi onkin hyvä, että Suomessa tupakointi asunnoissa on suurimmaksi osaksi kielletty. Ongelmat ovat kuitenkin laajempia kuin pelkkä sisätiloissa tupakointi. Muutkin tekijät, kuten yleinen siisteys, kosteus ja lämpötila, vaikuttavat sisäilman laatuun. Olennaista on myös aika, jonka ihminen viettää sisätiloissa. Tutkimusten mukaan ihminen käyttää yli 85 % ajastaan sisätiloissa (Spengler et al. 2001; Kosonen & Tan 2004; Ouaret et al. 2018).

Etenkin lapset ja nuoret viettävät paljon aikaansa sisällä käydessään koulua. Maltalla tehdyn tutkimuksen mukaan koululaisilla on suuri riski sairastua hengitystiesairauksiin, kuten astmaan, jos koulurakennuksen sisäilma on huono. Koulun sisäilma vaikuttaa suoraan koululaisten terveyteen. (Fsdani et al. 2018) Ervasti et al. (2012) tehdyn tutkimuksen mukaan huono sisäilma voi ennustaa myös opettajien ja muun koulujen henkilökunnan sairaspöissaoloja. Tutkimuksessa selvitettiin myös psykososiaalisten tekijöiden vaikutusta pöissaoloihin, mutta merkittävää eroa ei havaittu, vaikka tekijät vaikuttavatkin opettajien pöissaoloihin. Tuloksista käy ilmi, että kolmen vuoden tarkkailuajana lyhyen ajan sairaspöissaolot nousivat 25 %:sta 35 %:iin ja pitkän aikavälin pöissaolot 17 %:sta 28 %:iin huonon ilmanlaadun ympäristöissä. (Ervasti et al. 2012) Kymmenen ja yhden-toista prosenttiyksikön nousut lyhyissä ja pitkissä sairaspöissaoloissa ovat liikaa kolmen vuoden tarkkailuajalla.

Ervasti et al. (2012) huomasivat myös, että toisessa kyselyssä vastanneiden määrä oli pienentynyt, minkä tutkijat olettavat johtuvan terveydellisistä syistä. Joissain kouluissa ilmanlaatu oli parantunut ja pöissaolot vähentyneet. Tutkijat tulkitsivat pöissaolojen vähentymisen johtuvan ilmanlaadun parantumisesta. He tulivat lopputulokseen, että ilmanlaadun paranemisen myötä pöissaolot vähenevät, jolloin myös pöissaolojen kustannukset pienenevät. (Ervasti et al. 2012) Jos kouluissa on havaittu näin selkeitä korrelaatioita ilmanlaadun ja terveyden välillä, tulokset voidaan yleistää myös muihin työpaikkoihin, varsinkin toimistotyöhön.

Ilman tunkkaisuuden vaikutukset ulottuvat laajalle. Reijulan ja Sundman-Digertin (2004) toteuttaman työpaikkakyselyn mukaan yli kolmannes vastanneista (n = 11 154) piti sisäilmaa liian kuivana tai tunkkaisena, neljännes piti ympäristöä pölyisenä ja viidennes

ilmoitti kokevansa oireita sisäilman vuoksi. Kyselyyn vastasi yli 11 000 henkilöä 122:sta eri työpaikasta, joten kysely on ollut tarpeeksi laaja yleistämistä varten. Tätä asiaa olisi-kin syytä tutkia laajemmin, jotta ongelma saisi lisää näkyvyyttä ja siihen alettaisiin kiinnittää enemmän huomiota.

3.3 Sisäilman laadun heikkenemisen taloudelliset vaikutukset

Terveydellisten haittojen lisäksi on tutkittu myös työn tehokkuuden heikkenemistä. Spenglerin et al. (2001) mukaan ympäristö vaikuttaa työn tehokkuuteen jopa 20 %. Sisäilmasta johtuvien ongelmien vuoksi he laskivat tuottavuuden heikkenemiseksi 2 %. (Spengler et al. 2001) Kosonen ja Tan (2004) tutkivat ilmastoinnin vaikutusta tuottavuuteen, ja toteavat tutkimuksessaan pienimmän ilmavirran periaatteella ilmastoiduissa rakennuksissa tuottavuuden laskevan aina 5-13 %. He ovat myös laskeneet, että jopa yhden prosentin tuottavuuden nousu toimistotyöntekijän keskipalkalla säästää saman verran rahaa kuin ilmastoinnin parantaminen kustantaisi. (Kosonen & Tan 2004) Tämä tarkoittaa siis sitä, että työnantajalla tai kiinteistönhoitajalla ei ole mitään syytä pitää ilmastointia pienemmällä kuin hyvä työympäristö vaatii.

Näiden kaikkien ihmisten sairaspoissaoloista ja tehokkuuden laskemisesta tulee paljon ylimääräisiä kustannuksia kaikille. Varsinkin valtion tai kunnan viroissa työskentelevien poissaolot maksavat turhaan veronmaksajille silloin, jos ne lisääntyvät huonon sisäilman vuoksi. Yhdysvalloissa 90-luvulla toimistotyöntekijät ovat kattaneet noin puolet maan bruttokansantuotteesta (Traynor et al. 1993, Spengler et al. 2001 mukaan). Suomessa on muiden länsimaiden tapaan samanlaisia palveluita ja teollisuutta, joten suunnilleen samat luvut pätevät täälläkin. Suomen bruttokansantuote on noin 240 miljardia euroa, joista toimistotyöntekijät siis tuottavat noin 120 miljardia euroa (International Monetary Fund 2019). Oletetaan esimerkin vuoksi, että lyhyen aikavälin sairaspoissaolo on kaksi työpäivää ja käytetään aiemmin mainittua 25 %:ia poissaolojen määrään. Oletetaan myös, että kaikista Suomen 2,5:stä miljoonasta työntekijöistä noin kaksi miljoonaa on töissä sisätiloissa, jossa ilmanlaatu vaikuttaa hyvinvointiin (Tilastokeskus 2019). Kahdesta miljoonasta siis viisisataatuhatta työntekijää joutuu olemaan poissa töistä sairauden vuoksi ainakin kaksi päivää vuodessa. Esimerkiksi teknologiateollisuudessa 40 viikotuntia tekevällä on vuodessa 213 työpäivää, joista kahden päivän poissaolo vastaa lähes prosentin osuutta (Teknologiateollisuus 2019). Prosentti toimistotyöntekijöiden tuottamasta bruttokansantuotteesta on siis noin 1,2 miljardia euroa ja poissaolot töistä tarkoittavat menetettyjä tuloja.

Poissaolot tulevat kalliiksi myös työnantajille. Tilastojen mukaan toimistotyöntekijän keskimääräinen kuukausipalkka on noin 2500 euroa (Palkkadata 2019). Kuukaudessa on

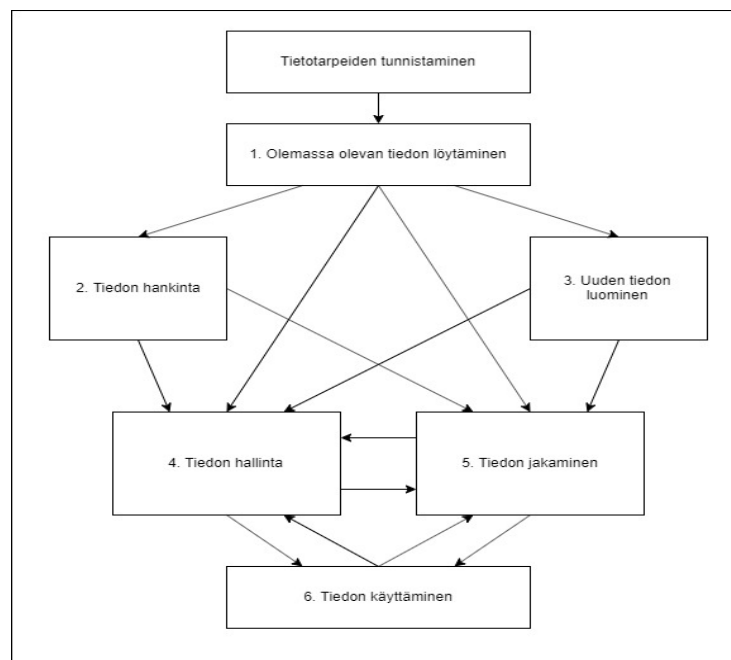
noin 25 työpäivää, jolloin kaksi sairaspäivää tarkoittaa 200:n euron turhaa kulua kuukaudessa per henkilö. Karkeasti laskettuna työnantajat siis menettävät viidensadantuhannen työntekijän poissaoloista 1,2 miljardia euroa sisäilman laadusta johtuvien ongelmien vuoksi. Sisätilojen viihtyvyydellä on suuria vaikutuksia, vaikka prosentti tuottavuudessa vaikuttaisikin mitättömältä. On siis järkevää miettiä tapoja, joilla saisi pienennettyä näitä turhia kuluja. Ylimääräisiä kuluja tulee myös rakennusten uusimisesta. Jos koulurakennus täytyy korjata kosteusvaurion vuoksi, oppilaat tulee ensin siirtää väliaikaistiloihin, minkä jälkeen vanha rakennus joko korjataan tai puretaan ja rakennetaan uusi. Tässäkin prosessissa tulee kustannuksia sen jokaisessa vaiheessa.

4. DATAN KÄSITTELY

Tietojohdamiseen liittyvän kirjallisuuden ja tutkimusten määrä on ollut kasvussa jo vuodesta 1995 lähtien (Bouthillier & Shearer 2002). Kirjoittaja on myös huomannut opintojen ja kesätöiden puolesta tarpeen liiketoiminnan kehittämiseksi tiedon ja informaation kautta. Laihosen et al. (2013) mukaan liiketoimintaa pyritään johtamaan tiedolla, jota saadaan jalostamalla dataa erilaisten toimintojen kautta. Tässä luvussa kerrotaan enemmän kyseisestä prosessista.

4.1 Analytiikkaprosessi

Analytiikkaprosessi on olennainen osa organisaatioiden informaatio- ja tietoprosesseja, jotka koostuvat seitsemästä askeleesta, joita ovat tietotarpeiden tunnistaminen, tiedon tai informaation hankinta, informaation siirto, informaation tallentaminen, tiedon tai informaation jakaminen, tiedon luominen ja tiedon tai informaation käyttö. Nämä prosessit mahdollistavat uuden tiedon luomisen ja organisaation kehitystoiminnan. (Bouthillier & Shearer 2002; Käsänpää 2017) Kuten kuvasta 3 näkyy, analytiikkaprosessin eri vaiheet kulkevat rinnakkain.



Kuva 3. Tiedolla johtamisen prosessi (Bouthillier & Shearer 2002 mukailen)

Wong et al. mukaan data-analytiikkaprosessi perustuu tiedonhankintaprosessille ja tiedonlouhinnan standardiprosessille (Wong et al. 2019). Rivon et al. (2012) mukaan tie-

donloughinnasta on tullut tärkeä menetelmä datan muuntamiseen tiedoksi. Heidän mukaansa tiedonloughinnalla tarkoitetaan kaavojen tunnistamista suurista datamääristä. (Rivo et al. 2012) Tiedonhankintaprosessissa etsitään tietoa ja informaatiota tietokannoissa olevasta datasta, ja se koostuu viidestä kohdasta, joita ovat datan valinta, datan esikäsittely, datan muuntaminen tiedoksi, tiedonloughinta, datan tulkinta sekä arviointi (Wang et al. 2019). Kokonaisdata koostuu useista pienemmistä datajoukoista, eli data-seteistä. Datasetti tarkoittaa samassa paikassa, esimerkiksi yhdessä tiedostossa tai kansiossa olevaa dataa (Crowder & Carbone 2018).

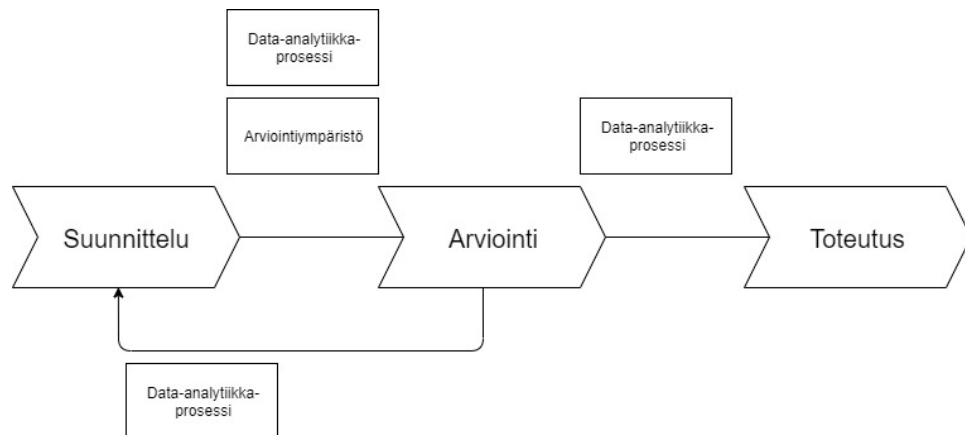
Monet tutkijat ovat ottaneet kantaa tiedonhankinnan prosesseihin, ja myös tiedonloughinnan standardiprosessi (cross-industry standard process of data mining) CRISP-DM oli alun perin EU:n tukema useiden yritysten toteuttama projekti. CRISP-DM jakaa tiedonhankintaprosessin kuuteen osaan (Rivo et al. 2012; Hiruta et al. 2019; Wang et al. 2019):

1. Liiketoiminnan ymmärtäminen: Tarkoituksena ymmärtää tavoitteet ja vaatimukset ensimmäisessä mallissa ja ongelman esittämisessä tiedonloughinnan näkökulmasta.
2. Datan ymmärtäminen: Alkuperäisen datasetin ja sen alaryhmien tutkimista ja niihin tutustumista erilaisista näkökulmista.
3. Datan valmistelu: Koko tiedonloughinnan prosessi, kuten datasetin jakamista alaryhmiin, datan järjestelyä ja datan muuntamista.
4. Mallinnus: Datan käsittelyyn ja mallintamiseen käytettävien työkalujen ja toimintojen valinta tiedonhankinnan kannalta.
5. Arviointi: Yhden tai useamman käytetyn mallin arviointia ensimmäisen kohdan kriteerien kannalta.
6. Jalkauttaminen: Mallin jalkauttaminen datan käsittelyyn piilossa olevan tiedon ja informaation löytämiseksi.

CRISP-DM koostuu helposti ymmärrettävistä kohdista, joita ei ole sidottu millekään alalle, joten se on siis yleisesti mille tahansa alalle sovellettava menetelmä datan käsittelyyn ja analysointiin.

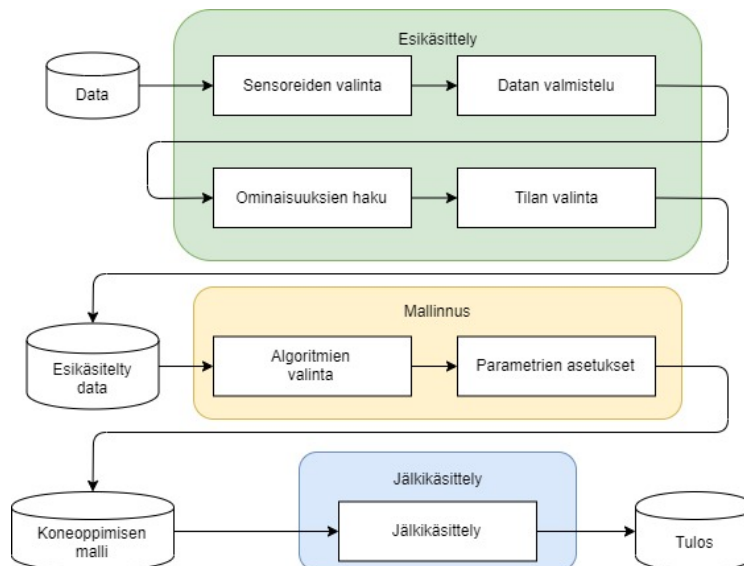
Hirutan et al. (2019) mukaan analytiikkaprosesseja ei kuitenkaan aina suunnitella samalla tavalla, sillä ongelmat ovat erilaisia. Esimerkiksi ongelmana voi olla analytiikan ja käytännön yhteensovittaminen, johon he esittivät ratkaisua koneoppimisen avulla. Esimerkkinä käytetään olosuhteiden seurantasysteemiä, johon analyttikot suunnittelivat analytiikkaprosessin.

Heidän analytiikkaprosessin suunnitteluprosessinsa koostuu kolmesta vaiheesta: suunnittelu, arviointi ja toteutus, kuten kuvasta 4 nähdään.



Kuva 4. Olosuhteiden seurantasysteemin analytiikkaprosessi (Hiruta et al. 2019 mukaillen)

Suunnitteluvaiheessa suunnitellaan prosessi yhdessä analyysin tulosten käyttäjien kanssa. Arviointivaiheessa analytikot arvioivat suunnittelun tuloksia arviointiympäristössä ja hyvässä tapauksessa prosessi siirretään toteutusvaiheeseen, jossa systeemi toteuttaa analytiikkaprosessia säännöllisesti. (Hiruta et al. 2019) Suunnittelu käyttäjien kanssa on tärkeää, sillä useat projektit epäonnistuvat siksi, että loppukäyttäjää kuunnellaan hyvin vähän tai ei ollenkaan.

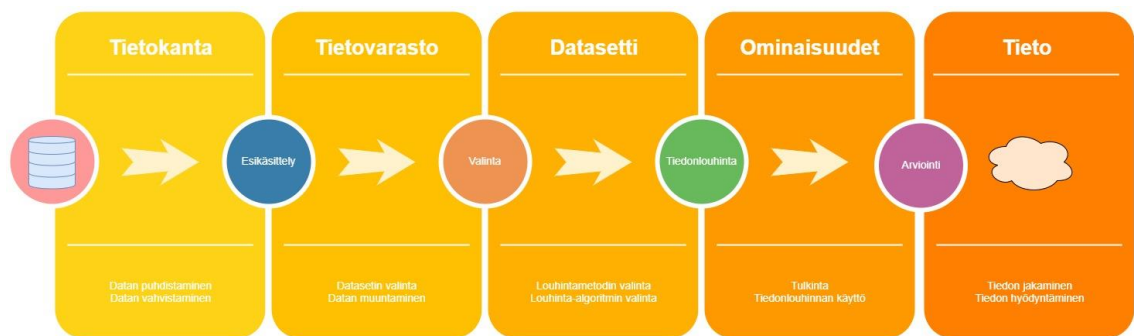


Kuva 5. Yleinen analytiikkaprosessi (Hiruta et al. 2019 mukaillen)

Hiruta et al. (2019) ovat luoneet olosuhteiden seurantasysteemin analytiikkaprosessin suunnittelijoiden kanssa käytyjen keskustelun pohjalta yleisen analytiikkaprosessin, joka koostuu esikäsittelystä, mallinnuksesta ja jälkikäsittelystä.

Kuten kuvasta 5 huomataan, analytiikkaprosessilla on paljon samaa kuin CRISP-DM -mallissa. CRISP-DM on vielä kuitenkin yleisempi ja siihen liittyy myös liiketoiminnallisia ominaisuuksia. Tämä prosessi on tarkennettu nimenomaan datan käsittelyyn ja mallin-
nus haluttiin tehdä koneoppimisen malleihin.

Wang et al. (2019) ovat kehittäneet CRISP-DM -menetelmän pohjalta oman viiden as-
keleen analytiikkaprosessin, joka koostuu tietokannasta, tietovarastosta, datasetistä,
ominaisuuksista ja tiedosta. Kuvan 6 mukaan askelten väliin vaaditaan erilaisia toimin-
toja, jotta voidaan siirtyä seuraavaan kohtaan. Heidän analytiikkaprosessinsa sopii eri-
tyisesti suurten datamäärien käsittelyyn. (Wang et al. 2019)



Kuva 6. Analytiikkaprosessi (Wang et al. 2019 mukaillen)

Näiden mallien pohjalta on pohdittu prosessin kokonaiskuvaa, jonka voisi esittää vielä
selkeämmin. Työn tutkimuksen kannalta oleellinen prosessi on esitelty kuvassa 7. Bout-
hillierin ja Shearerin (2002) mallin tavoin myös tässä prosessissa käsitellään asioita rin-
nakkain, vaikka nuoli onkin vain yhteen suuntaan.



Kuva 7. Analytiikkaprosessi yksinkertaistettuna

Yleensä tiedon tarve johtuu jostakin ongelmasta tai tilanteesta, jossa ei osata tehdä pää-
töstä tilanteen ratkaisemiseksi. Analytiikkaprosessi kuvaa toimintoja tietotarpeiden ha-
vaitsemisesta tarpeiden täyttämiseen vaadittaviin päätöksiin.

4.2 Koneoppiminen

Wein et al. (2019) mukaan analytiikkaa varten on kehitetty erilaisia työkaluja, mutta koneoppiminen (engl. machine learning) on niistä yleisin. Koneoppiminen perustuu ohjattuun (engl. supervised) ja ohjaamattomaan (engl. unsupervised) oppimiseen. Ohjatun oppimisen algoritmit käyttävät valmista dataa harjoittamaan algoritmeja tunnistamaan tuntemattomat datapisteet. Ohjaamaton oppiminen puolestaan käyttää tuntematonta dataa harjoittamaan algoritmeja analysoimaan ja tunnistamaan datan rakenteen, jotta sitä voidaan käyttää ohjatun oppimisen algoritmeissa tekemään ennustavia malleja. (Wei et al. 2019) Koneoppiminen perustuu neuroverkoihin, ja siihen liittyy usein kaavojen tunnistamista ja signaalinkäsittelyä.

Neuroverkot puolestaan perustuvat neuronien tai yksiköiden keskinäisille yhteyksille, jotka käyttävät painotettuja funktioita muuntamaan syötearvoja ennustettaviin arvoihin ilman, että tarvitaan erillistä oletusta muuttujien välisistä yhteyksistä. Neuroverkkoon kuuluu kolme peruskerrosta: syötekerros, piilotettu kerros ja tuotekerros. Piilotettuja kerroksia voi olla useita. Neuroverkkoja voi olla useita erilaisia, mutta yleisin tapa on tehdä useita piilotettuja kerroksia ja harjoittaa dataa paljon samalla, kun funktioiden ja neuronien painotuksia päivitetään edellisten tulosten perusteella. (Wei et al. 2019) Neuroverkkojen käyttöön liittyy paljon teknistä osaamista, johon on syytä tutustua lisää.

5. SISÄILMAN LAADUN KÄSITTELY DATA-ANALYTIIKAN MENETELMIN

Tässä luvussa vastataan tutkimuskysymykseen: Miten ilmanlaadun heikkenemistä voidaan ehkäistä ennen hyvinvoinnin heikkenemistä? Aluksi esitellään ilmanlaadun mittaamiseen liittyvän analytiikkaprosessi, jonka jälkeen pohditaan mitä analyysin tuloksilla voi tehdä tavoitteen saavuttamiseksi.

5.1 Datan käsittely ja analysointi

Kuten analytiikkaprosessissa todettiin, on olemassa ongelma tai tilanne, jonka ratkaisemiseksi vaaditaan lisää tietoa. Useissa tutkimuksissa on todettu, että huono sisäilman laatu aiheuttaa hengitystiesairauksia ja syöpää (Hoskins 2003; Zhang & Smith 2003; Ervasti et al. 2012; Fsdani et al. 2018). Ihmisten sairastaminen heikentää elämänlaatua ja aiheuttaa ylimääräisiä kustannuksia, mikä puolestaan nopeuttaa elämänlaadun heikkenemisen kierrettä. Elämänlaatua halutaan parantaa, jolloin on havaittu tarve elämänlaatua parantavalle tiedolle.

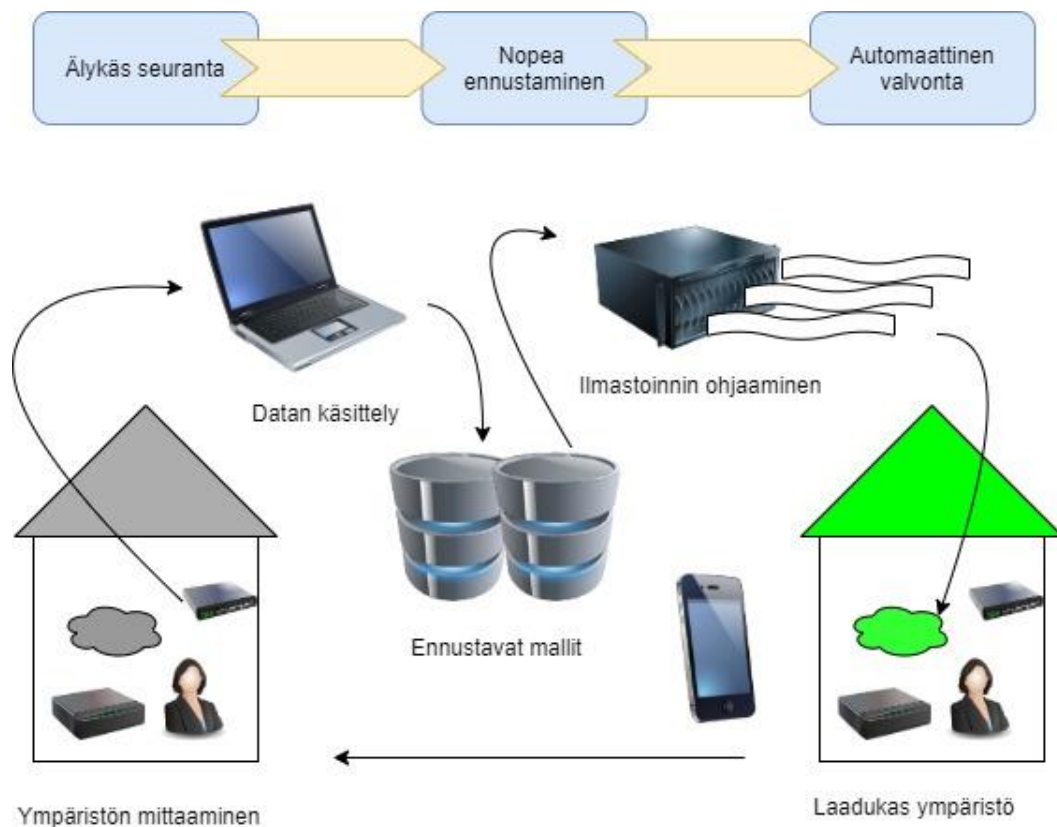
Esimerkiksi juuri koneoppiminen ja neuroverkot sopivat ilmanlaadun mittaamisesta saadun datan käsittelyyn. Caronin et al. (2019) mukaan mitattu data voidaan luokitella eri yhdisteiksi K-means –klusteroinnilla. Ranganathanin (2013) mukaan klusterointi on joukkojen muodostamista samankaltaisista alkioista, ja k-means on yleisin koneoppimisen klusterointialgoritmi. Siinä datasetistä erotellaan haluttu määrä ryhmiä, joihin muodostetaan keskustoja, joiden ympärille muodostetaan datasta ryhmät. Tarvittaessa keskustoja muodostetaan lisää tai niiden määrää vähennetään. Algoritmia iteroidaan niin kauan kuin on tarve, jotta saadaan realistinen kuva datasta. (Ranganathan 2013)

Neuroverkkojen harjoittamista varten täytyy valita harjoitusdataa, jonka koko voi olla esimerkiksi 60 % koko datasetistä. K-means –klusteroinnin avulla tunnistetut yhdisteet syötetään algoritmille, minkä jälkeen neuroverkkoja harjoitetaan harjoitusdatan avulla. (Saad et al. 2017) Neuroverkot kehittävät mallin, jota voidaan käyttää ennustamaan eri yhdisteiden pitoisuuksia tietyllä ajanjaksolla.

Näitä malleja täytyy kuitenkin tarkastella kriittisesti ja arvioida onnistumista prosessin jokaisessa vaiheessa. Putra et al. (2018) käyttävät neuroverkkojen arviointiprosessissa kahta kohtaa. Ensin prosessia arvioidaan keskineliövirheen (engl. mean square error) avulla. Keskineliövirhe mittaa tulosten poikkeamien neliöiden keskiarvot. Keskineliövirhe

ilmenee datasetin satunnaisuuden tai puuttuvan tiedon vuoksi. Siksi mitataan myös regressiota, eli muuttujien keskinäistä suhdetta. (Putra et al. 2018)

Neuroverkkojen ja koneoppimisen avulla saadaan tietoa ilmanlaadun tilasta. Esimerkiksi tilan hiilidioksidipitoisuus kasvaa luonnollisesti silloin, kun ihmisten määrä kasvaa. Samanlaiset yhteydet pystytään selvittämään myös muille vaikeammin havaittaville yhdisteille, kuten typen yhdisteille. Pidemmällä aikavälillä myös voidaan nähdä mikäli tilan il-
mankosteus tai lämpötila nousee, jolloin voidaan epäillä kosteusvaurion kehittymistä ja tehdä tarvittavia toimenpiteitä tilanteen pahenemisen estämiseksi.

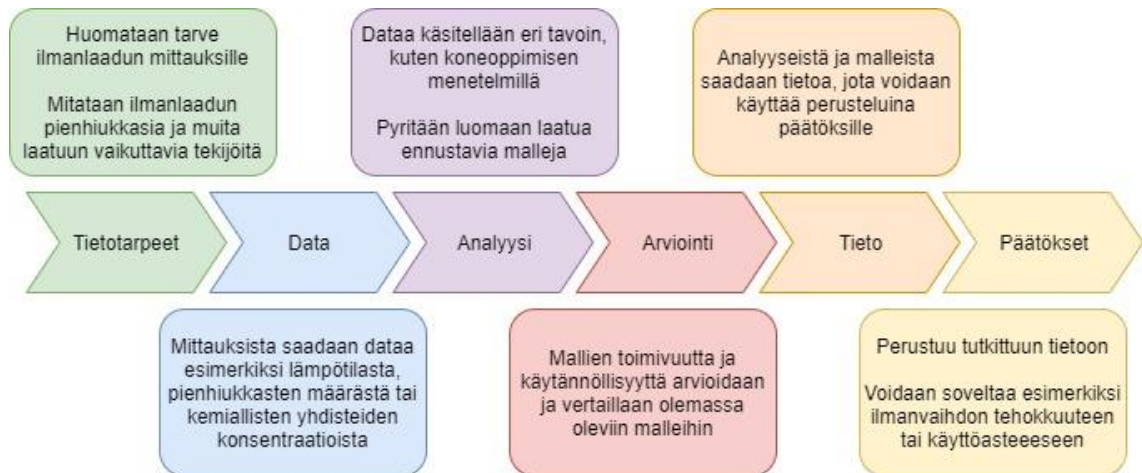


Kuva 8. Reaaliaikainen ilmanlaadun seurantajärjestelmä (Ren & Cao 2020 mukailen)

Esimerkiksi ilmanlaatusensorit voivat lähettää dataa reaaliajassa ilmastointilaitteille, tai aikaisemmin mitatun datan perusteella laitteet voidaan ohjata toimimaan siihen aikaan, jolloin ilmastoinnin tarve on suuri. Kuvassa 8 esitellään automatisoidun ilmanlaadun seurantajärjestelmän toimintaperiaatetta. Mittareista saatavaa dataa käsitellään neuroverkkojen avulla ja ennustetaan tuleva ilmastoinnin tarve, jolloin elinympäristö pysyy viihtyisenä jatkuvasti. Renin ja Caon (2020) mukaan reaaliaikaa nopeampi reagointi ilmanlaatuun on tavoiteltava tila siksi, että se on tehokkaampaa energian käyttöä. Lisäksi

aiemmin mainitut terveydelliset ja taloudelliset syyt tekevät ilmanlaadun mittaamisesta tärkeää.

Neljännessä luvussa esitelty analytiikkaprosessi voidaankin näiden perusteiden pohjalta esittää myös ilmanlaadun mittaamisen kontekstissa. Prosessia voidaan käyttää esimerkiksi ilmastoinnin suunnittelua ja käyttöä koskevien päätösten tekemisen tukena.



Kuva 9. Ilmanlaadun mittausten analytiikkaprosessi

Kuvassa 9 on kuvattu neljännessä luvussa esitelty analytiikkaprosessi sisäilman laadun seurannan kontekstissa. Prosessin jokaista kohtaa on avattu ja kerrottu, mitä sisäilman laadun analytiikkaprosessissa tehdään. Tämä prosessi on tässä kandidaatintyössä olennainen, ja vastaa työn päätutkimuskysymykseen.

5.2 Tiedolla johtaminen

Talojen rakentaminen ei tapahdu hetkessä, eikä niistä päättä yksi ihminen. Siksi on tärkeää, että rakenteiden, materiaalien ja ilmastoinnin suunnittelulle on vakaat perustelut. Ilmanlaatumittareista saatava data on peruste tiedolle, jota voidaan käyttää rakennussuunnittelussa. Ghasemaghaei (2019) mukaan datan on kuitenkin oltava laadukasta, jotta siitä on hyötyä yritysjohton tekemissä päätöksissä. Myös data-analytiikan on oltava laadukasta, jotta siitä on hyötyä. Huono analytiikan käyttö voi johtaa negatiivisiin tuloksiin, jos datamäärät ovat suuria. Analyysimenetelmillä ei juurikaan ollut vaikutusta tuloksiin datamäärien ollessa pieniä. (Ghasemaghaei 2019)

Data-analytiikan lisäksi tärkeä elementti organisaation johdon päätöksiin vaikuttamisessa on datan visualisointi. Mooren (2017) mukaan datan esittämisen tulee olla yksinkertaista, informatiivista ja selkeää. Esityksessä on otettava huomioon se, kenelle esitetään, ja mitä he tietävät valmiiksi. Hyvässä esityksessä annetaan myös analytiikan nä-

kemys päätöksestä, mitä datan perusteella kannattaisi tehdä. (Moore 2017) Jotkut ymmärtävät paremmin kuvaajia ja taulukoita kuin toiset. Ilmanlaatuakin voi välillä joutua esittämään hymynaamoilla, jotta analytiikkaan perehtymättömätkin tietävät, milloin laatu on hyvää tai huonoa.

Eräs sisäilman laadun mittauksiin erikoistunut yritys tuottaa mittaus- ja analytiikkapalveluitaan kiinteistövälitysyriyksille. Kiinteistön ostajaa kiinnostaa rakennuksen kunto, jolloin sieltä mitataan lämpötilaa, kosteutta ja orgaanisia yhdisteitä. Samanlaista prosessia voitaisiin soveltaa myös julkisella puolella, esimerkiksi rakennusten huoltojen suunnittelussa. Julkisella puolella haasteita tuovat kuitenkin rahoituksen puute sekä poliittiset näkemykset varojen käyttämisestä.

Mooren (2017) ja Ghasemaghain (2019) mukaan analytiikka on kuitenkin tärkeä osa päätöksentekoa moderneissa organisaatioissa. Kaikkialla kerätään dataa, mutta sitä ei aina osata hyödyntää oikein. Analyttisin menetelmin ja oikein esitettynä datasta saadaan informaatiota ja tietoa, jonka pohjalta voidaan tehdä päätöksiä terveyden, talouden ja kestäväen kehityksen edistämiseksi. (Moore 2017; Ghasemaghain 2019)

6. YHTEENVETO

Viimeisessä luvussa esitellään ja arvioidaan tutkimuksen onnistumisia ja haasteita. Lopuksi myös arvioidaan tulosten merkitystä ja kerrotaan, miten aihetta kannattaa tutkia tulevaisuudessa.

6.1 Tulosten esittely ja arviointi

Tutkimus eteni suoraviivaisesti ja järkevästi taustasta ja menetelmistä kohti lopputulosta. Aiheen rajaus oli järkevä ja siitä saatiin paljon tietoa monipuolisista lähteistä. Erityisesti tutkimuksen taustasta löytyi runsaasti tietoa, ja välillä piti etsiä pitkään rajauksen mukaisia lähteitä. Sisäilman laadun vaikutuksista kertova kirjallisuus oli odotetusti vanhempaa kuin muut, mutta sekin oli edelleen relevanttia ja laadukasta. Sisäilman laadusta ja sen vaikutuksista löytyneet tutkimukset olivat selkeitä. Yllättävää oli, että suuri osa tutkimuksista oli suomalaisia tai suomalaisten toteuttamia (Kosonen & Tan 2004; Reijula & Sundman-Digert 2004; Tampio 2010; Ervasti et al. 2012; Finell et al. 2018). Tästä voidaankin päätellä, että Suomessa asia on edelleen ajankohtainen ja siihen kiinnitetään jo valmiiksi enemmän huomiota kuin muualla maailmassa. Kuitenkaan se, että kiinnitetään enemmän huomiota kuin aiemmin ei tarkoita sitä, että aihetta olisi tutkittu tarpeeksi. Aihetta on siis syytä tutkia lisää.

Analytiikkaprosessista löytyneet lähteet olivat hieman uudempia, sillä data-analytiikan soveltaminen liiketoiminnassa on vielä suhteellisen uusi käsite. Analytiikasta kertovaa kirjallisuutta piti etsiä vielä tarkemmin kuin ilmanlaadun vaikutuksia. Löytyneessä kirjallisuudessa analytiikkaa sovellettiin usein johonkin muuhun kuin mitä etsittiin, jolloin aikaa piti käyttää enemmän aiheeseen sopivan kirjallisuuden etsimiseen. Analytiikkaprosessista ja ilmanlaadun seurannasta kuitenkin löytyi tietoa, joka auttoi tutkimaan aihetta lisää. Tutkimuskysymyksiin ei löytynyt suoraa vastausta, jolloin eri lähteistä löytynyttä tietoa piti soveltaa ja yhdistää oikeaan kontekstiin. Tulokset saatiin tutkimalla sisäilman laadun tekijöitä ja vaikutuksia sekä sen mittaamista ja analysointia. Tutkimuksen tuloksena luotu prosessi sisäilman laadun mittauksiin on suoraviivainen ja johdonmukainen.

Sisäilman laadun mittauksia varten luotu analytiikkaprosessi koostuu kuudesta kohdasta: tietotarpeet, data, analyysi, arviointi, tieto ja päätökset. Prosessi alkaa tietotarpeiden tunnistamisella, minkä jälkeen tietoa täytyy kerätä mittaamalla ilmanlaatua. Mittauksista saadaan dataa, jota tulee analysoida esimerkiksi koneoppimisen menetelmillä.

Analyyseistä luodaan ennustavia malleja, joita arvioidaan ja vertaillaan jo olemassa oleviin malleihin. Analyyseistä ja malleista saadaan tietoa, jota voidaan käyttää perusteluina päätöksille. Päätökset perustuvat siis tutkittuun tietoon ja tuloksia voidaan soveltaa esimerkiksi ilmanvaihdon tehokkuuteen tai käyttöasteeseen.

Tutkimuksen aihe oli kirjoittajalle tuttu, mutta syvempää osaamista aiheeseen ei ollut. Tutkimuksessa pyrittiin etsimään ja luomaan uutta tietoa, joka osoittautui paikoitellen hankalaksi. Uutta tietoa löytyi paljon uusista julkaisuista, joista jopa yksi artikkeli julkaistaan vasta tämän työn valmistumisen jälkeen. Uutta tietoa oli kuitenkin haasteellista luoda, sillä ilmanlaadun analyttistä seuranta on tutkittu paljon. Uutuusarvoa tuo toivottavasti analyttisten menetelmien erilainen muotoilu ja yksinkertaisempi esittäminen, sekä niiden yhdistäminen ilmanlaadun mittauksiin ja tiedolla johtamisen konseptiin.

Tutkimuksessa ei ilmennyt muita merkittäviä haasteita. Aikaa vei kuitenkin aiheen rajaaminen, sillä aineistoa oli todella paljon saatavilla. Osittain siitäkin syystä tekniseen puoleen perehtyminen jäi vähäiseksi. Työn laajuus on suunniteltu olevan noin 20-30 sivua ja aikaa kirjoittamiseen oli noin seitsemän viikkoa. Siinä ajassa aiheeseen perehtyminen syvällisemmin oli haasteellista, mutta rajauksen puitteissa työn tekeminen oli tehty melko helpoksi.

6.2 Tutkimuksen merkityksellisyys ja lisätutkimusten tarve

Ilmanlaadun mittaamisella on todettu olevan monia hyötyjä, eikä ollenkaan haittoja. Siksi mittaamisen tärkeyttä tulisi painottaa entistä enemmän. Mittaamalla sisäilman laatua saadaan dataa, jota analysoimalla voidaan arvioida tulevaa energiankäyttöä sekä säästää ympäristöä ja rahaa. Kontrolloimalla sisäilman laatua voidaan myös säästää ihmisten terveyttä ja tehdä työtä tehokkaammin. Suurin merkitys sisäilman laadun mittauksille on rakennustekniikan puolella. Taloja remontoidaan joskus jopa täysin turhaan, kun yksinkertaisilla ja edullisilla mittauksilla voisi selvittää mahdollisuuden ongelman eristämiseen tai pienempiin korjauksiin. Tutkimuksen tavoitteena on, että sisäilman laadun mittaukset yleistyisivät, jotta ihmisten hyvinvointi paranisi. Sillä välin aihetta tulisi tutkia lisää.

Työssä on tutkittu mittaamista ja analysointia sekä todettu, että tällaiselle toiminnalle on olemassa oikea tarve. Ilmanlaadun mittaamista ja analysointia varten täytyisi kuitenkin olla selkeät ohjeet, jotta prosessin aloittaminen olisi selkeää ja helppoa. Tämän työn tarkoitus oli selvittää yleisiä prosesseja ja menetelmiä datan hyödyntämiseksi. Kandidaattityön laajuus ei kuitenkaan riitä tutkimaan asiaa perinpohjaisesti, joten siksi aihetta tulisi tutkia lisää teknisen osaamisen ja siitä saatavan ymmärryksen kautta. Yleisesti myös

koneoppimisen menetelmiä ja neuroverkkoja kannattaa tutkia lisää, jotta niiden käyttö yleistyisi ja olisi helpompaa.

LÄHTEET

- Ben-David, T., Rackes, A., Lo, L., Wen, J. & Waring, M. (2019). Optimizing ventilation: Theoretical study on increasing rates in offices to maximize occupant productivity with constrained additional energy use. *Building and Environment*. Vol. 166, pp. 118-133.
- Berkeley Lab. (2019a). Overview of IAQ. Saatavissa: <https://iaqscience.lbl.gov/overview> (2.10.2019)
- Berkeley Lab. (2019b). Perceived Indoor Air Quality and Performance. Saatavissa: <https://iaqscience.lbl.gov/performance-perceived> (2.10.2019)
- Bouthillier, F. & Shearer, K. (2002). Understanding knowledge management and information management: the need for an empirical perspective. *Information Research*. Vol. 8(1).
- Caron, A., Redon, N., Coddeville, P. & Hanoune, B. (2019). Identification of indoor air quality events using a K-means clustering analysis of gas sensors data. *Sensors and Actuators B: Chemical*. Vol. 297.
- Crowder, J. & Carbone, J. (2018). Data Analytics: The Big Data Analytics Process (BDAP) Architecture. Proceedings of the International Conference on Information and Knowledge Engineering (IKE). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). pp. 118-123.
- Daisey, J., Angell, W. & Apte, M. (2003). Indoor air quality, ventilation and health symptoms in schools: an analysis of existing information. *Indoor Air*. Vol. 13(1). pp 53-64.
- Ervasti, J., Kivimäki, M., Kawachi, I., Subramanian, S., Pentti, J., Oksanen, T., Puusniekka, R., Pohjonen, T., Vahtera, J. & Virtanen, M. (2012). School environment as predictor of teacher sick leave: data-linked prospective cohort study. *BMC Public Health*. Vol. 12(1). p. 770.
- Finell, E., Tolvanen, A., Pekkanen, J., Minkkinen, J., Ståhl & T., Rimpelä, A. (2018). Psychosocial Problems, Indoor Air-Related Symptoms, and Perceived Indoor Air Quality among Students in Schools without Indoor Air Problems: A Longitudinal Study. *International Journal of Environmental Research and Public Health*. Vol. 15(7). p. 1497.
- Fsdani, P., Bezzina, F., Fsdani, C. & Montefort, S. (2018). Impact of School Air Quality on Children's Respiratory Health. *Indian Journal of Occupational & Environmental Medicine*. Vol. 22(3). pp. 156-162.
- Ghasemaghaei, M. (2019). Does data analytics use improve firm decision making quality? The role of knowledge sharing and data analytics competency. *Decision Support Systems*. Vol. 120. pp. 14-24.
- Hiruta, T., Uchida, T., Yuda, S. & Umeda, Y. (2019). A design method of data analytics process for condition based maintenance. *CIRP Annals*. Vol. 68(1). pp. 145-148.
- Hoskins, J. (2003). Health Effects due to Indoor Air Pollution. *Indoor and Built Environment*. Vol. 12(6). pp. 427-433.

- International Monetary Fund. (2019) World Economic Outlook Database. Saatavissa: https://www.imf.org/external/pubs/ft/weo/2019/02/weo-data/weorept.aspx?sy=2017&ey=2022&scsm=1&ssd=1&sort=coun-try&ds=.&br=1&pr1.x=78&pr1.y=13&c=172&s=NGDP_R%2CNGDP_RPCH%2CNGDP%2CNGDPPC%2CLP&grp=0&a= (4.11.2019)
- Kosonen, R. & Tan, F. (2004). The effect of perceived indoor air quality on productivity loss. *Energy and Buildings*. Vol. 36(10). pp. 981-986.
- Känsänkoski, H. (2017). Informaatio- ja tietoprosessit tietojohdamisen viitekehyksenä terveydenhuollon organisaatioissa – kohti potilasarvon luomista. *Informaatiotutkimus*. Vol. 36(1).
- Laihonen, H., Hannula, M., Helander, N., Ilvonen, I., Jussila, J., Kukko, M., Kärkkäinen, H., Lönnqvist, A., Myllärniemi, J. & Pekkola, S. (2013). Tietojohdaminen, Tampereen teknillinen yliopisto - tiedonhallinnan ja logistiikan laitos, Tampere.
- Langer, S., Ramalho, O., Le Ponner, E., Derbez, M., Kirchner, S. & Mandin, C. (2017). Perceived indoor air quality and its relationship to air pollutants in French dwellings. *Indoor Air*. Vol. 27(6). p. 1168-1176.
- Moore, J. (2017). Data Visualization in Support of Executive Decision Making. *Interdisciplinary Journal of Information, Knowledge & Management*. Vol. 12. pp.125-138.
- Ouaret, R., Ionescu, A., Petrehus, V., Canday, Y. & Ramahlho, O. (2018). Spectral band decomposition combined with nonlinear models: application to indoor formaldehyde concentration forecasting. *Stochastic Environmental Research and Risk Assessment*. Vol. 32(4). pp. 985-997.
- Palkkadata. (2019). Hallinto- ja toimistotehtävät. Saatavissa: <https://www.palkkadata.fi/salaryinfo/hallinto-ja-toimistotehtavat/toimistotyontekija?search=1>. (4.11.2019)
- Putra, J.C.P., Safrilah & Ihsan, M. (2018). The prediction of indoor air quality in office room using artificial neural network. Presented at the HUMAN-DEDICATED SUSTAINABLE PRODUCT AND PROCESS DESIGN: MATERIALS, RESOURCES, AND ENERGY: Proceedings of the 4th International Conference on Engineering, Technology, and Industrial Application (ICETIA) 2017, Surakarta, Indonesia.
- Ranganathan, S. (2013). Improvements to k-means clustering. *Diplomityö*. Tampereen teknillinen yliopisto. p. 47.
- Ren, C. & Cao, S. (2020). Implementation and visualization of artificial intelligent ventilation control system using fast prediction models and limited monitoring data. *Sustainable Cities and Society*. Vol. 52(1).
- Reijula, K. & Sundman-Digert, C. (2004). Assessment of indoor air problems at work with a questionnaire. *Occupational & Environmental Medicine*. Vol. 61(1). pp. 33-38.
- Rivo, E., de la Fuente, J., Rivo, Á., García-Fontán, E., Cañizares, M. & Gil, P. (2012). Cross-Industry Standard Process for data mining is applicable to the lung cancer surgery domain, improving decision making as well as knowledge and quality management. *Clinical and Translational Oncology*. Vol. 14(1). pp. 73-79.

Saad, S. M., Allan, M. A., Shakaff, A. Y., Dzahir, M. A., Hussein, M., Mohamad, M. & Ahmad, Z. A. (2017). Pollutant Recognition Based on Supervised Machine Learning for Indoor Air Quality Monitoring Systems. *Applied Sciences*. Vol. 7(8). p. 823.

Salminen, A. 2011. Mikä kirjallisuuskatsaus? Johdatus kirjallisuuskatsauksen tyypeihin ja hallintotieteellisiin sovelluksiin. Vaasan yliopisto, Vaasa.

Spengler, J., Samet, M. & McCarthy, J. (2001). *Indoor Air Quality Handbook*. McGraw-Hill Education. Saatavilla: <https://www.accessengineeringlibrary.com/content/book/9780074455494>. Luettu: 29.10.

Tampio, E. 2010. Ilmanlaadun seurannan kehittämistarpeet. Pro gradu -tutkielma. Jyväskylän yliopisto. 95 s.

Teknolohiateollisuus. (2019). Työntekijöiden työaika 2017-2019. Saatavissa: https://teknolohiateollisuus.fi/sites/default/files/file_attachments/vuosityoaika2017-2019_40_375.pdf. (4.11.2019)

Terveyden ja hyvinvoinnin laitos, 2019. Mitkä tekijät vaikuttavat sisäilman laatuun? Saatavissa: <http://thl.fi/fi/web/ymparistoterveys/sisailma/mitka-tekijat-vaikuttavat-sisailman-laatuun->. Luettu: 9.10.

Tilastokeskus. (2019). Työvoimatutkimus. Saatavissa: https://www.stat.fi/til/tyti/2018/13/tyti_2018_13_2019-04-11_kat_002_fi.html. (4.11.2019)

Wang, C., Lin, S., Chou, T. & Li, B. (2019). An integrated data analytics process to optimize data governance of non-profit organization. *Computers in Human Behavior*. Vol. 101(12). pp.495-505.

Wei, W., Ramalho, O., Malingre, L., Sivanantham, S., Little, J. C. & Mandin, C. (2019). Machine learning and statistical models for predicting indoor air quality. *Indoor Air*. Vol. 29(5). pp. 704-726.

Zhang, J. & Smith, K. (2003). Indoor air pollution: a global health concern. *British Medical Bulletin*. Vol. 68(1). pp. 209-225.